Construction d'images miniatures avec recadrage automatique basée sur un modèle perceptuel bio-inspiré

Olivier Le Meur, Patrick Le Callet, Dominique Barba

Construction d'images miniatures avec recadrage automatique basée sur un modèle perceptuel bio-inspiré

O. Le Meur[1] *P. Le Callet*[2] *D. Barba*[2]

[1] *THOMSON*
1, Avenue de Belle-Fontaine - BP 19
35511 Cesson-Sévigné
France

[2] *IRCCyN*
Rue Christian Pauc - BP 50609
44306 Nantes
France

Résumé

Cet article présente un procédé de zoom automatique, destiné à adapter la taille des images pour des dispositifs d'affichage à écran de petite taille (Téléphone mobile...). L'adaptation de la taille des images s'effectue par la sélection des zones les plus intéressantes visuellement. Ces dernières sont déterminées via une approche computationnelle de modélisation de l'attention visuelle. Tout d'abord, les performances de modélisation de l'attention visuelle sont déduites par comparaison avec une vérité terrain issue de tests oculométriques. Enfin, les performances qualitatives du système de vision complet, modèle d'attention visuelle et procédé de miniaturisation, sont présentés.

Mots clefs

attention visuelle exogène, système visuel, images miniatures.

Abstract

This article presents an automatic process used to automatically crop images. This tool is necessary to meet the constraints (small display) of the new portable device (Mobile phone..). The principle of the reframing rests on both the detection and the selection on the most interesting parts of the picture. These areas are detected by a computational model of the bottom-up visual attention. First, the results of this model are compared to results stemming from an eye tracking apparatus. Second, the performances of the whole system are presented.

Key words

Bottom-up visual attention, visual system, eye tracking experiments, thumbnail.

1 Introduction

Bien que l'environnement visuel dans lequel nous évoluons soit constitué d'une quantité considérable d'information, notre système visuel est capable d'appréhender et d'interpréter avec précision l'ensemble de ces informations visuelles. Des mécanismes particuliers ainsi que des stratégies d'exploration de l'espace visuel sont nécessaires pour résoudre cette situation paradoxale. Les premiers permettent de construire une représentation économique du contenu visuel. Cette représentation, où la redondance a été supprimée, est précise au centre de la rétine (la fovéa) et grossière dans la

périphérie. Concernant les stratégies d'explorations, elles sont de deux types: l'attention endogène et l'attention exogène. La première, également appelée *Top-Down*, est une stratégie pilotée par la tâche que nous effectuons à un instant donné, impliquant un contrôle volontaire et cognitif des mouvements oculaires. Ce mécanisme, nécessitant toutes les ressources attentionnelles, est déployé pour effectuer une tâche: reconnaître un lieu sur une photo, chercher l'homme portant une casquette verte... La stratégie exogène, plus communément appelée *Bottom-Up*, permet, quant à elle, de sélectionner les informations visuelles selon leur degré de saillance. Ce type de stratégie fait référence à l'attention visuelle involontaire, c'est à dire à un traitement automatique très rapide réalisé inconsciemment. Le déplacement du regard est dans ce cas piloté par les caractéristiques spatio-temporelles du flux lumineux.

Dans cette contribution, un système de vision constitué d'un modèle d'attention visuelle est proposé. La modélisation de l'attention visuelle est dans le système envisagé basée uniquement sur les caractéristiques bas niveaux du signal (luminance et couleur). En d'autres termes, les aspects cognitifs, les informations dites de haut niveau (le texte par exemple) ne sont pas pris en compte. Ce système a pour objectif d'effectuer un cadrage et un zoom automatique construit à partir de l'extraction des zones perceptuellement intéressantes. Ce système de vision peut être utilisé pour diverses applications: l'une des plus importantes concerne le domaine de la robotique (pilotage d'un robot de façon automatique). Une autre concerne les images et la mobilité. La profusion de nouveaux dispositifs portables, caractérisés par un écran de faible taille, nécessite d'adapter les contenus transmis à ces nouvelles contraintes d'affichage. Pour les terminaux ayant un écran de faible taille, un dispositif de création d'images miniatures ou encore vignettes (*thumbnail* en anglais) est nécessaire. Comme présenté ultérieurement, il existe deux façons de construire des images miniatures: réduire la taille par simple décimation ou sélectionner une sous-partie de l'image à partir d'un modèle d'attention visuelle et l'agrandir à la taille désirée. La première est celle classiquement utilisée. Elle consiste à décimer ou sous-échantillonner l'image d'entrée afin de prendre en compte les contraintes d'affichage. La seconde est tout à fait différente. Contrairement à l'approche de décimation, la totalité de l'image n'est pas affichée. Seule une partie, regroupant les zones perceptuellement intéressantes est affichée.

A partir d'un modèle bio-inspiré de la partie *Bottom-Up* du mécanisme d'attention visuelle chez l'humain, modèle que nous avons développé et présenté dernièrement [Le Meur 06], cette contribution décrit un système dédié à la miniaturisation d'images, faisant intervenir successivement la modélisation de l'attention visuelle et un procédé de construction d'images de tailles réduites. Le modèle d'attention visuelle que nous avons utilisé est donc brièvement introduit dans la première partie, le lecteur pouvant se reporter aux références [Le Meur 05], [Le Meur 06] pour plus de détails. Cette modélisation est essentiellement de type bas niveaux dans le sens où ce sont les parties antérieures du système visuel humain qui sont pris en compte (rétine, aire V1 et aire V2 du cortex visuel). Elle ne prend pas en compte un système de reconnaissance des formes, même partiel dans ses capacités à reconnaître, qui pourrait alors nous permettre de générer une saillance cognitive, sûrement plus proche de la saillance visuelle que produit un humain. Ensuite, l'évaluation quantitative du modèle est réalisée à partir d'une nouvelle façon de calculer une vérité terrain provenant de tests occulométriques. Dans nos tests préliminaires publiés dans [Le Meur 06], la vérité terrain prenait en compte la durée de fixation, ce qui n'est plus le cas ici. Enfin, le système de création d'images miniatures avec recadrage automatique est détaillé et ses performances évaluées via un protocole spécifique de tests.

2 Modèle d'attention visuelle sur images fixes

2.1 Principe général

La plupart des modèles psycho-visuels se basent sur l'architecture biologiquement plausible proposée par C. Koch et S. Ullman [Koch 85]. Cette architecture s'inspire clairement des propriétés du système visuel ainsi que de son fonctionnement. Elle consiste à effectuer un premier traitement automatique et massivement parallèle sur l'ensemble du champ visuel. Cette étape est suivie d'un traitement localisé, permettant d'extraire les caractéristiques visuellement les plus marquantes.
Le premier traitement extrait des caractéristiques visuelles de la scène et les regroupe dans un certain nombre de cartes. Chaque carte fournit une représentation déformée mais topologiquement exacte du champ visuel. Bien qu'il n'y ait pas de listes exhaustives des caractéristiques à extraire, les plus courantes sont l'orientation, la couleur, la courbure, le changement de régularité...
Le second traitement fait référence à la localisation des éléments qui "sautent aux yeux". C. Koch et S. Ullman définissent la saillance, c'est à dire le degré d'attractivité d'un site, comme une fonction dépendante de l'inhibition latérale entre cellules[1]. En d'autres termes, dans une carte d'une caractéristique visuelle donnée, la saillance d'une zone dépend de son contexte local.
Le problème majeur de cette architecture concerne l'extraction des caractéristiques visuelles. Chaque caractéristique visuelle nécessite un procédé d'extraction propre, produisant des paramètres donnés. Il y a donc potentiellement autant de procédés d'extraction que de caractéristiques visuelles, pouvant générer des dynamiques fondamentalement différentes. La comparaison et la combinaison de cartes provenant de différentes caractéristiques visuelles s'avèrent donc très délicates.

La modélisation que nous proposons s'appuie également sur l'architecture de C. Koch et S. Ullman, mais la philosophie sous-jacente est différente. A partir d'une image incidente, un espace psycho-visuel est construit. Il est constitué des composantes naturelles de notre environnement, c'est à dire d'une composante achromatique et de deux composantes chromatiques. La différence fondamentale vis-à-vis de l'état de l'art se situe dans la normalisation de ces composantes. En effet, afin d'obtenir des caractéristiques visuelles homogènes et donc comparables, elles sont toutes normalisées par rapport à leur seuil différentiel de visibilité propre. Ainsi, que ce soit une valeur liée à une composante achromatique ou à une composante chromatique, elles s'expriment toujours en fonction de leur seuil de visibilité. Comme évoqué précédemment, le premier intérêt est d'avoir des données homogènes, donc comparables. Le deuxième concerne la hiérarchisation des données. Les informations inférieures au seuil de visibilité sont négligées alors que d'autres sont mises en exergue. On ne s'intéresse donc qu'aux données perceptibles par le système visuel et donc susceptible d'attirer le regard.
A partir de ce cadre conceptuellement cohérent, la mesure de saillance de chaque site reste à déterminer. Il s'agit de transformer des valeurs exprimées en terme de visibilité en valeurs de saillance. Pour cela, il est nécessaire de bien faire la différence entre visibilité et saillance.
La visibilité caractérise l'état d'un stimulus qui peut être visible ou invisible. Comme le passage d'un état à l'autre n'est pas instantané, la visibilité d'un stimulus est mesurée grâce à des expériences psychophysiques. Ces dernières cherchent à déterminer la probabilité de détection d'un stimulus lorsque les paramètres de ce dernier (amplitude, orientation...) varient. Concernant la saillance et bien qu'il n'y ait aucune définition formelle, le terme de saillance se réfère à l'intérêt, à l'attention ou encore la priorité qu'un site porte. La saillance qualifie également l'attractivité visuelle d'un site et par conséquent le rôle qu'il est susceptible de jouer dans la stratégie visuelle.
Dans cette modélisation et à partir de l'espace psycho-visuel, une carte de saillance pour chaque composante (une composante achromatique et deux composantes chromatiques) est extraite.

[1] On considère ici uniquement la saillance physique. D'autres types de saillance existent tels que la saillance cognitive, qui dépend des processus cognitifs de l'observateur traitant le message. Pour une revue exhaustive des différents types de saillance, le lecteur pourra se reporter à [Landragin 04].

2.2 Espace de visibilité

L'obtention de l'espace de visibilité (ou encore espace psycho-visuel) est illustrée à la figure 1.

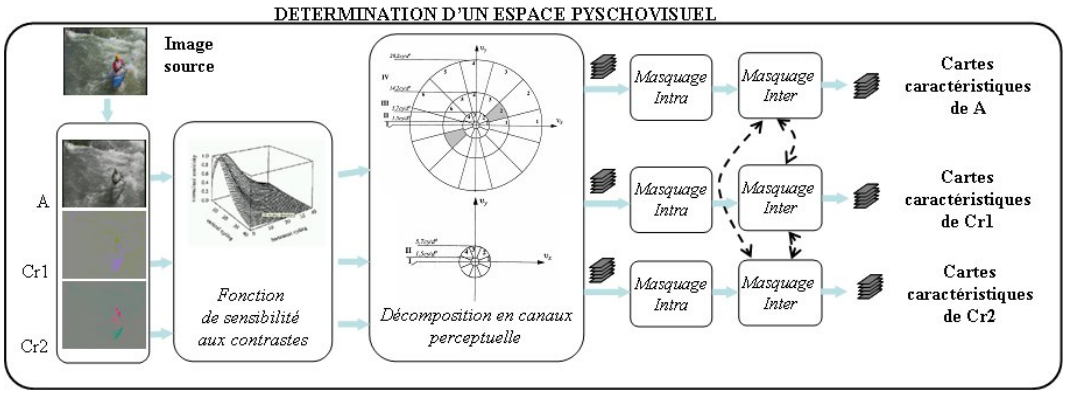


Fig. 1: Mécanisme d'obtention de l'espace psycho-visuel.

L'espace de visibilité est constitué des fonctions brièvement rappelées ici (le lecteur pourra se référer à [Le Meur 06] pour de plus amples détails):

- Espace perceptuel de représentation couleur: le système visuel décompose l'information lumineuse incidente en trois composantes distinctes, une achromatique, notée A et deux chromatiques, notées Cr_1 et Cr_2. L'espace de représentation choisi est celui défini par J. Krauskopf [Krauskopf 82];

- Fonction de sensibilité aux contrastes : la sensibilité du système visuel à une stimulation visuelle dépend de nombreux paramètres. Des expériences psychophysiques permettent de maîtriser un certain nombre d'entre eux (fréquences spatiales, orientations, luminances ambiantes). Des fonctions de sensibilité au contraste (CSF, *Contrast Sensitivity Functions*) anisotropiques permettent de considérer les deux premiers paramètres. Il est clair que plus la fréquence spatiale est élevée et plus la sensibilité du système visuel est faible. Concernant l'orientation, la sensibilité aux composantes horizontales et verticales est plus forte que la sensibilité aux composantes diagonales; le système visuel humain, confronté à un environnement visuel majoritairement composé de composantes horizontales et verticales, a développé une grande sensibilité à ces orientations. Chaque composante fréquentielle appartenant aux canaux (A, Cr_1, Cr_2) est donc pondérée par une CSF;

- Décomposition en sous-bandes perceptuelles : étant donné que les cellules du système visuel réagissent à des stimulations particulières, il est fondamental de considérer le système visuel comme un procédé décomposant l'information visuelle en un ensemble de canaux, décrits en termes de sélectivité en fréquences radiales et en termes de sélectivité angulaire. Chaque canal, ou sous bande visuelle, représente une population de cellules répondant à un type de stimulus particulier. La première transformation simulant les différentes populations de cellules fut proposée par A. Watson, avec la transformée Cortex [Watson 87]. Des études approfondies, faisant appel à de nombreuses expérimentations psychophysiques menées lors de différents travaux [Senane 93] et [Le Callet 99], ont permis de redéfinir les paramètres de cette transformée. La composante A est décomposée en 17 sous-bandes réparties sur 4 couronnes (un canal basses fréquences non directionnel (noté I) et trois bandes de fréquences radiales directionnelles (notées de II à IV). Les couronnes II à IV sont décomposées en canaux angulaires dont le nombre varie avec la bande de fréquences radiales considérée;

- Masquage visuel : la décomposition en sous-bandes perceptuelles permet de simuler le pavage fréquentiel du système visuel. Bien que les cellules visuelles liées à une sous-bande donnée réagissent fortement à une stimulation particulière, la réponse d'une cellule dépend également de la présence d'autres stimuli, c'est à dire du contexte dans lequel elle se trouve. Ce phénomène est appelé effet de masquage, dans le sens où, généralement, cela induit une augmentation du seuil différentiel de visibilité. Comme les précédentes fonctions, il est détaillé et référencé dans [Le Meur 06].

2.3 Contraste de visibilité

En simulant les comportements intrinsèques des cellules visuelles, l'espace de visibilité a permis de présélectionner et de hiérarchiser implicitement les informations visuelles. Néanmoins, la quantité d'information à traiter reste considérable. Il s'agit donc maintenant d'extraire les caractéristiques visuelles portant de l'information importante. Pour cela, les valeurs de visibilité des données de l'espace psycho-visuel sont transformées en saillance. Rappelons que la saillance est une fonction dépendante de l'inhibition latérale entre cellules. Ainsi, être singleton, c'est à dire être en contraste avec son environnement local, est le facteur prépondérant qui génère de la saillance. Un exemple type, (contraste de couleur) concerne le fait qu'une tache d'une couleur C1 (rouge par exemple) mélangée au sein d'une population de taches d'une autre couleur C2 très différente de C1 (verte par exemple) attire d'emblée le regard alors que la population de taches de l'autre couleur n'est perçue que dans son ensemble [Wolfe 04].
Dans le système visuel humain, cette détection de contraste est rendue possible par la structuration des champs récepteurs des cellules visuelles, que ce soit au niveau rétinien ou cortical. Ces champs récepteurs sont radiaux, soit de type circulaire pour les cellules dans la rétine, soit de type ellipsoïdal avec une direction préférée pour les cellules corticales et sont constitués d'un centre et d'un pourtour ayant des réponses antagonistes. Cette organisation leur confère donc la propriété de répondre fortement sur les contrastes et de ne pas répondre sur les zones uniformes. La modélisation de ce type de cellules s'effectue via des différences de Gaussiennes (DoG) orientées ou non.
Ainsi, les sous-bandes provenant des trois composantes sont convoluées avec un opérateur proche d'une DoG, générant une représentation dite économique du champ visuel. Pour des raisons de complexité calculatoire, la détection des contrastes de visibilité est réalisée uniquement sur la couronne III, car cette dernière représente les fréquences spatiales intermédiaires où la sensibilité visuelle est maximale. Elle présente par ailleurs un bon compromis entre quantité d'information et précision de la localisation.

2.4 Carte de saillance par canal et carte de saillance finale

2.4.1 Carte de saillance achromatique et chromatique

Trois cartes de saillance sont déterminées: une carte de saillance achromatique et deux cartes de saillance chromatique. Elles sont obtenues de la façon suivante. Pour la carte de saillance achromatique, en sommant simplement les sous-bandes orientées de la couronne III, couvrant la bande de fréquences radiales (5.7 - 14.2 cycles par degré) car c'est la couronne dans laquelle le système visuel humain est le plus sensible (car excitée par les zones de contour et de texture). Pour les deux cartes de saillance chromatique, on opère de même mais sur la couronne II, la couronne III n'étant pas exploitée dans notre décomposition en sous-bandes visuelles. C'est par soucis de simplification que nous nous restreignions par composante (A, Cr1, Cr2) à une seule bande de fréquences radiales. Les autres bandes radiales interviennent également dans le système visuel humain mais sont de moindre importance. De plus, des pondérations, favorisant telles ou telles orientations, peuvent être utilisées. Ce type de pondération est ici implicitement effectué par les CSFs anisotropes.

2.4.2 Carte de saillance finale

Finalement, afin de construire la carte de saillance spatiale finale, une fusion des différentes composantes est effectuée groupant ou liant des éléments, a priori indépendants, pour former une structure compréhensible par le cerveau. La fusion est basée sur deux mécanismes:

- une compétition intra carte permettant d'identifier les zones les plus pertinentes des cartes;
- une compétition inter cartes tirant profit de la redondance et de la complémentarité des différentes cartes de saillance. L'utilisation de la redondance inter cartes permet de renforcer la saillance de certaines zones lorsque celles-ci génèrent de la saillance dans plusieurs dimensions. Par contre, lorsqu'une zone ne génère de la saillance que dans une seule dimension visuelle, il est nécessaire d'utiliser la complémentarité inter cartes.

Ce mécanisme de fusion est entièrement décrit dans [Le Meur 05].

3 Performances

Afin d'évaluer les performances du modèle d'attention visuelle, il est nécessaire de disposer d'une vérité terrain. Cette dernière est déterminée à partir d'expérimentations oculométriques. Il s'agit ensuite d'évaluer le degré de similarité entre les données prédites et expérimentales.

3.1 Expérimentations oculométriques

Un dispositif oculomètrique est un dispositif permettant de suivre les déplacements de la pupille. Ce type de dispositif est basé sur la capacité de nos yeux à réfléchir les infrarouges. En fait, deux types de reflets sont observés: des reflets fixes dus à la réflexion des infrarouges sur la cornée (reflets de Purkinje), et des reflets mobiles dus à la réflexion des infrarouges sur la pupille. La position relative de ces deux types de reflets permet de déterminer le positionnement de l'oeil. Les données recueillies peuvent être exploitées de différentes façons. Les zones attirant le regard des observateurs sont déterminées. La durée de fixation est un élément intéressant pour mesurer le degré de saillance d'une zone[2]. Par ailleurs, la stratégie visuelle, c'est à dire le déplacement oculaire, peut faire l'objet d'étude, même si cela semble a priori difficile à aborder du fait de l'idiosyncrasie de la stratégie visuelle.
L'expérimentation débute par une étape de calibrage. Elle consiste à afficher un certain nombre de points à l'écran et à enregistrer les mouvements oculaires. La correspondance entre position du regard et position des points permet de déterminer la relation existante entre les mouvements oculaires et la sortie de l'oculomètre. Pendant le test, si nécessaire, une nouvelle étape de calibrage peut être réalisée. L'expérimentation commence par l'affichage des images de test. L'écran utilisé était un CRT de résolution 800×600. La taille de l'écran était de 36×27 cm et la distance de visualisation de 81 cm. Les images de test sont au nombre de 24 et présentent des contenus variés (paysage, sport, visage, scène d'intérieur et d'extérieur). Elles sont affichées pendant une durée de 15 secondes (pour la construction de la vérité terrain, on ne considère que les 8 premières secondes de visualisation).
Le tableau 1 récapitule les données importantes du protocole expérimental pour les tests oculométriques sur images fixes.

Chaque image de test est observée en visualisation libre. La seule instruction donnée aux observateurs est de regarder l'image le plus naturellement possible. Pour chaque image et pour tous les sujets, les points de fixation forment un ensemble PF, contenant autant d'éléments que de fixations. Un point de fixation k, noté f_k se caractérise par ces coordonnées (x_k, y_k). Un point

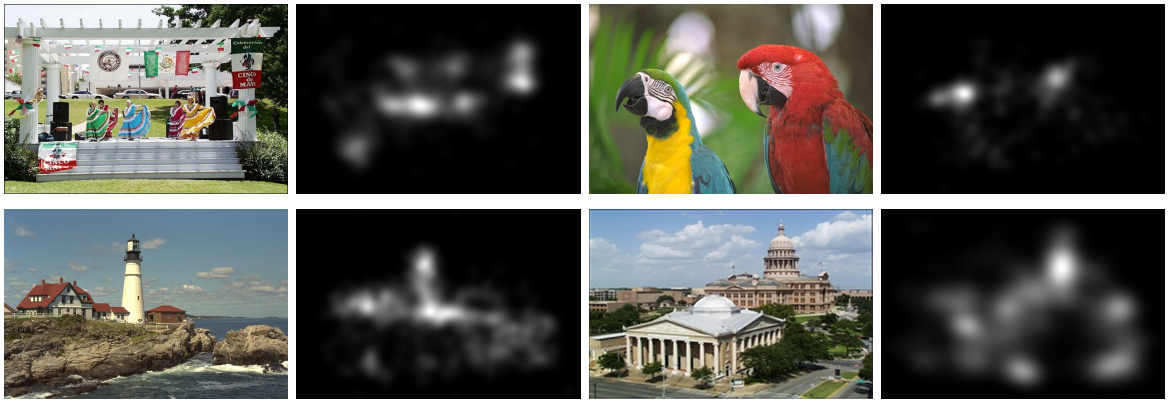
[2] La durée de fixation est un élément important pour mesurer le degré de saillance. Néanmoins, la saillance n'est pas directement proportionnelle à la durée de fixation [Itti 05].

Tab. 1: Protocole des expérimentations pour l'acquisition de données oculométriques sur images fixes.

Protocole	
Distance d'observation	$3H$
	H hauteur de l'écran
Résolution de l'écran	800×600
Nombre d'images traitées	40
Type d'images	niveaux de gris et couleur
Nombre d'observateurs	40
Durée de l'observation	$15s$
Calibrage	20 points de calibrage

(a) Première fixation (b) Deuxième fixation (c) Troisième fixation (d) Quatrième fixation

Fig. 2: Positions spatiales des points de fixation. Chaque carte représente respectivement le premier (a), le deuxième (b), le troisième (c) et le quatrième point de fixation pour 12 observateurs.



(a) Image source (b) Carte de saillance (c) Image source (d) Carte de saillance

Fig. 3: Cartes de saillance issues des expérimentations oculométriques: (a) et (c) images sources; (b) et (d) cartes de saillance.

de fixation doit durer plus de 100 ms. Les saccades et les fixations de courtes durées ne sont pas considérées. Cette façon de procéder diffère radicalement des précédentes évaluations effectuées [Le Meur 06], pour lesquelles la durée de fixation était prise en compte[3].
La carte d'attention finale pour une image donnée, notée SM, est déduite de l'ensemble PF:

$$SM(x,y) = (\sum_{k=1}^{N}(\Delta(x-x_k, y-y_k))) * g_\sigma(x,y) \quad (1)$$

où, N est le nombre de fixation, (x_k, y_k) les coordonnées spatiales discrètes du point de fixation f_k et g_σ le noyau de convolution d'un filtre 2D gaussien discret. Le symbole de Kronecker Δ est défini par :$\Delta(n,m) = 1$ si $(n,m) = (0,0)$, 0 sinon. L'écart-type σ est déterminé en fonction de la précision de la mesure de l'oculomètre (0.5 degré d'angle visuel). Par ailleurs, ce type de filtrage prend également en compte le fait qu'un observateur fixe une zone spatiale plutôt qu'un point.

La figure 2 présente les premiers points de fixation pour différents observateurs. La figure 3, quant à elle, donne des exemples de cartes de saillance. Les zones blanches représentent les zones attirant le plus le regard des observateurs.

3.2 Evaluations quantitatives

Deux métriques, le coefficient de corrélation linéaire et la divergence de Kullback-Leibler, sont utilisées pour évaluer le degré de similarité existant entre les cartes d'attention visuelles issues des expérimentations et celles prédites par le modèle.

3.2.1 Coefficient de corrélation

Le coefficient de corrélation linéaire, classiquement utilisé, mesure le degré de linéarité existant entre les deux fonctions spatiales de saillance (modèle et observateur moyen). Invariante par transformée linéaire et bornée dans l'intervalle $[0,1]$, cette mesure, notée cc, est donnée par la relation suivante:

$$cc(p,h) = \frac{cov(p,h)}{\sigma_p \sigma_h} \quad (2)$$

[3] Le lecteur ne pourra donc pas retrouver facilement une cohérence entre les différents résultats donnés. Par ailleurs, notons également que la technique de fusion présentée dans cet article, elle aussi, est radicalement différentes

Tab. 2: Comparaison de différents modèles. Les valeurs suivies du signe * sont statistiquement différentes de la valeur obtenue par le modèle proposée (student, $p < 0.05$).

Modèle	$cc \pm SEM$	$kl \pm SEM$
Modèle proposé	0.45 ± 0.07	1.58 ± 0.24
Modèle aléatoire	$0.0012 \pm 0.00004^*$	$13.23 \pm 0.59^*$
Modèle uniforme	0^*	$3.72 \pm 0.28^*$
Modèle centré	0.45 ± 0.03	$2.6 \pm 0.26^*$
Modèle proposé + pondération	$0.52 \pm 0.06^*$	$1.34 \pm 0.17^*$

où, h et p représentent, respectivement, les cartes de saillance visuelle expérimentale et prédite définies sur le même domaine spatial, $cov(p, h)$ représente la covariance entre p et h. σ_p et σ_h sont respectivement l'écart type de p et h.
Sur les 24 images testées, le coefficient de corrélation moyen est de 0.45 ± 0.07 ($cc \pm SEM$)[4].
Le tableau 2 présente les résultats de différents modèles testés: un modèle aléatoire, un modèle uniforme, un modèle centré (distance par rapport au centre de l'image) et le modèle proposé associé à une pondération centrée.

Tout d'abord, notons que le modèle aléatoire et le modèle uniforme ne sont pas en mesure de prédire la saillance. Le mode de sélection des zones observées par les observateurs n'est donc ni aléatoire ni uniforme. A l'inverse, le modèle proposé ainsi que le modèle centré sont les meilleurs descripteurs de la saillance. Le modèle centré mets en évidence l'importance du centre de l'image. Cette importance est due à un aspect cognitif et expérimental. Le premier concerne la prise de vue qui, dans de nombreux cas, place la zone visuellement intéressante au centre ou proche du centre de l'image. Le second concerne la vérité terrain déduite des tests oculométriques. Dans ces tests, l'oculométre est centré sur l'écran et les images à tester apparaissent centrées sur l'écran. Ces conditions de tests ne contribuent donc pas à atténuer le pouvoir attracteur du centre de l'image (elles auraient plutôt tendance à l'augmenter).
En prenant en compte cette dépendance au centre, c'est à dire en pondérant les cartes de saillance par une gaussienne centré sur l'image (son écart-type a été optimisé afin de fournir les meilleurs résultats [Le Meur 06]), le coefficient de corrélation passe de 0.45 à 0.52. Cette heuristique n'est valable que si on se place dans un contexte de visualisation d'images affichées sur un écran. Elle n'a pas de sens si il s'agit de modéliser l'attention visuelle de façon générale.
Il est important de mentionner les points suivants qui expliquent en partie pourquoi le coefficient de corrélation est relativement faible. Tout d'abord, il existe une différence fondamentale entre les cartes expérimentales et les cartes prédites. La distribution des premières est à bords francs avec un taux de couverture relativement faible. Le taux de couverture est le ratio entre le nombre de pixel ayant une valeur de saillance supérieure à un seuil et le nombre total de pixel. A titre d'exemple, une image, pour laquelle la saillance est diffuse (c'est souvent le cas des paysages), présentera un taux de couverture élevé. Par contre, une image contenant une zone contrastant fortement avec le fond présentera un taux de couverture plus faible (les observateurs auront tendance à focaliser d'avantage aux mêmes endroits). Par contre, la distribution des cartes prédites est plus lissée et présente un taux de couverture plus important. Par ailleurs, existe-t-il une relation linéaire entre la saillance expérimentale et la saillance prédite? En d'autres termes, une valeur de saillance expérimentale s est-elle équivalente à une valeur de saillance prédite $\alpha \times s$. Pour prendre en compte cette interrogation, des opérateurs non-linéaires (exponentielle, logarithmique, puissance...) ont été appliqués sur la carte de saillance prédite. L'application de ces pré-traitements n'a pas permis d'augmenter le coefficient moyen de corrélation linéaire. Dans le même ordre d'idée, une valeur de saillance expérimentale de 10 attire-t-elle 10 fois plus l'attention qu'une zone de saillance

[4] la SEM, pour *Standard Error of the Mean* est obtenue en divisant l'écart type par la racine carré de la taille n de l'échantillon testé. La SEM est utilisée pour calculer l'intervalle de confiance à 95%, $cc \pm \lambda \times SEM$ (Pour un échantillon de taille inférieur à 30, le facteur multiplicatif λ est déterminé à partir d'une loi de Student avec $n-1$ degré de liberté; si la taille de l'échantillon est supérieur à 30, $\lambda = 1.96$.)

expérimentale de 1? Enfin, bien que les tests oculométriques aient été réalisés en exploration libre, la saillance expérimentale est le résultat de la fusion des deux mécanismes attentionnels, c'est à dire l'attention exogène et l'attention endogène.

3.2.2 Divergence de Kullback-Leibler

La divergence de Kullback-Leibler, notée kl, évalue le degré de dissimilarité entre deux fonctions de probabilités. Elle est donnée par la relation suivante:

$$kl(p|h) = \sum_x p(x) Log(\frac{p(x)}{h(x)}) \tag{3}$$

où, h et p représentent respectivement les densités de probabilités associées aux cartes d'attention expérimentale et prédite. Lorsque les deux densités de probabilité sont strictement égales, la valeur kl est nulle. Cette métrique n'est pas symétrique et n'a pas de borne supérieure.
Sur les 24 images testées, la divergence de Kullback-Leibler est de 1.58 ± 0.24 (cf. tableau 2). Comparativement aux résultats précédents, on constate que le modèle centré ne donne pas de bons résultats. Les erreurs de prédictions commises par ce dernier sont souvent grossières: l'exemple de l'image des perroquets (première ligne de la figure 2) pour laquelle les objets saillants ne sont pas centrés est significatif ($kl = 4.08$ pour le modèle centré, $kl = 0.99$ pour le modèle proposé). La même constatation peut être faite pour l'image du phare (seconde ligne de la figure 2): $kl = 3.32$ pour le modèle centré, $kl = 1.09$ pour le modèle proposé.
Toutefois, l'application de la pondération sur le modèle proposé améliore de façon significative la valeur moyenne du degré de dissimilarité (il passe de 1.58 à 1.34).

4 Application à la création d'images miniatures

Cette application concerne la construction d'images miniatures facilitant à la fois la recherche de contenu et améliorant la satisfaction visuelle des utilisateurs. Face à l'explosion des contenus multimédia et aux offres pléthoriques de création de contenus, des outils d'aide à la recherche de contenus et permettant d'adapter un contenu à différentes tailles d'écran s'avèrent nécessaires. Typiquement, l'objectif est d'améliorer le confort visuel des dispositifs à écran de taille réduite (téléphone portable, PC de poche...). Dans ce contexte, l'idée est de sélectionner et d'afficher seulement les parties d'intérêt d'une image en effectuant un recadrage automatique et une mise à l'échelle.

4.1 Des travaux récents

Les premiers travaux dans ce domaine sont ceux de X. Fan et al. [Fan 03]. L'objectif était de faciliter l'exploration d'une base d'images affichées sur l'écran d'un téléphone portable. Les zones d'intérêt sont déterminées à partir d'une carte de saillance. Cette dernière, issue du modèle de L. Itti, est couplée avec des informations de haut niveau telles que la détection de texte et la détection de visage. La zone la plus saillante est alors extraite de l'image. Une amélioration de ce procédé a été très vite proposée [Liu 03]. Elle consiste à prendre en compte le fait qu'une image peut présenter plusieurs zones d'intérêt spatialement éloignées. Dans un tel contexte, lorsqu'un observateur regarde l'image, l'oeil effectue des séquences de fixations sur les zones d'intérêt et de saccades entre les différentes zones. Pour reproduire ce phénomènes, H. Liu et al. [Liu 03] proposent d'afficher séquentiellement chaque zone d'intérêt en fonction de leur saillance. Un paramètre, appelé MPT pour *Minimal Perceptible Time*, est intégré. La durée d'affichage est fonction de la saillance et des préférences de l'utilisateur.
Ces travaux, très séduisants sur certains aspects, ne mettent pas assez l'accent sur la nécessité d'adapter la taille de la miniature en fonction de la distribution de la saillance de l'image. Pour

pallier ce problème, nous proposons une solution entièrement automatique permettant d'adapter la taille de la miniature en fonction notamment de la distribution de la saillance.

4.2 Images miniatures centrées sur les zones visuellement intéressantes

La création d'images miniatures doit répondre à plusieurs exigences. La plus importante, à notre avis, concerne la taille de la miniature. Doit-elle être fixe et centrée sur le maximum global ou adaptée en fonction de certains paramètres tels que la distribution de la saillance? Si on considère une image ayant un taux de couverture faible, la taille de la miniature peut être petite favorisant un effet de zoom sur la zone d'intérêt. Par contre, si le taux de couverture est important, la miniature doit avoir une taille raisonnable pour englober la majorité des zones saillantes.

4.2.1 Sélection des sites les plus saillants

A partir de la carte de saillance finale, c'est à dire issue des composantes achromatiques et chromatiques, un algorithme de type *Winner-Take-All* est utilisé pour déterminer les maximums locaux. Le maximum local $n+1$ est déterminé en inhibant une zone circulaire $\mathcal{C}$ centrée autour du maximum local n et d'un rayon tel que la zone soit vue sous un angle de un degré visuel; cette taille de zone est en relation directe avec la taille du champ visuel fovéal. Chaque zone circulaire représente donc une partie intéressante de l'image. Cette zone circulaire, centrée sur un maximum, est repositionnée sur son centre d'inertie ci avant d'être inhibée. Ce dernier est calculé de la façon suivante:

$$ci_x = \frac{1}{\sum_{(x,y)\in\mathcal{C}} SM(x,y)} \sum_{(x,y)\in x\mathcal{C}} SM(x,y) \quad (4)$$

$$ci_y = \frac{1}{\sum_{(x,y)\in\mathcal{C}} SM(x,y)} \sum_{(x,y)\in y\mathcal{C}} SM(x,y) \quad (5)$$

avec, ci_x et ci_y les coordonnées du centre d'inertie ci. SM représente la carte de saillance spatiale finale.
La recherche des maximums locaux s'arrête lorsque la saillance inhibée contenue dans les cercles recentrés représente plus de $P\%$ de la saillance totale de l'image. La valeur de $P\%$ est importante car c'est elle qui conditionne la taille finale de la miniature. Cette valeur est adaptée en fonction du taux de couverture de l'image traitée. Rappelons que le taux de couverture qualifie la distribution de la saillance; une valeur faible tend à montrer qu'il existe une et une seule régions d'intérêt alors qu'une valeur forte signifie que l'image ne contient aucune zone vraiment saillante.

4.2.2 Construction de l'image miniature

A partir des coordonnées des coins des carrés dans lesquels sont inscrits les cercles centrés sur un maximum local de saillance, un sous-ensemble est déterminé. Les coordonnées des coins en haut à gauche et en bas à droite de ce sous-ensemble sont déduites afin d'inclure toutes les zones carrés centrées sur des maximums locaux pertinents. La conservation du ratio ou la mise à ratio unitaire peut alors être effectuée (en fonction du besoin utilisateur). Le sous-ensemble est finalement remis à l'échelle pour être dimensionnée à la taille souhaitée de la miniature.
La figure 4 présente deux situations différentes, mettant en évidence l'importance de la détermination de la taille de la miniature; elle dépend du nombre de maximums choisi mais également de leurs positions spatiales relatives. La première situation concerne l'image *Kayak*. Cette image contient une et une seule zone saillante. Les différents maximums locaux pertinents se situent tous autour de cette zone saillante. Dans ce cas, la pertinence visuelle de la miniature est très élevée. La seconde situation concerne l'image *Perroquets* pour laquelle deux zones saillantes apparaissent clairement. L'image miniature obtenue présente une pertinence moyenne car elle intègre les deux zones visuellement intéressantes mais également d'autres zones moins intéressantes. Un ensemble de résultats est donné à la figure 5.

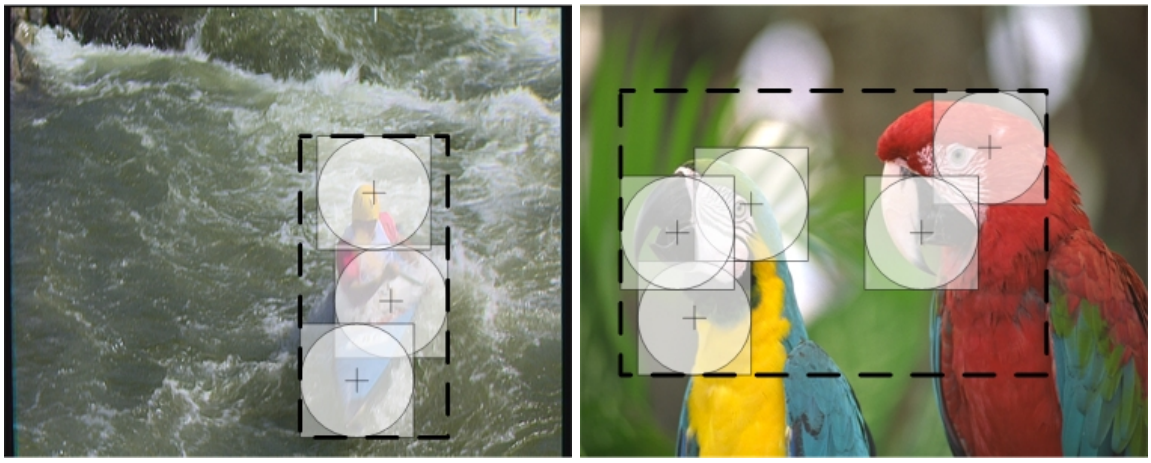


Fig. 4: Exemples de miniatures centrées sur les zone saillantes: (a) image *Kayak* avec trois maximums locaux et (b) image *Perroquets* avec cinq maximum locaux. L'image miniature résultante est représentée par la rectangle en pointillé noir.

4.2.3 Evaluation qualitative

Pour réaliser l'évaluation qualitative, on se place dans un contexte de recherche de contenus et de visualisation d'images de grandes tailles sur des écrans de tailles réduites. Typiquement, on se place dans la situation où un grand nombre d'images réduites sont affichées sur un écran; ces images formant en quelque sorte une mosaïque. Ces images aux formats réduits n'offrent pas toutes les mêmes conditions de visualisation. Certaines sont exploitables immédiatement, c'est à dire que, en un coup d'oeil, l'observateur est capable d'appréhender le contenu de l'image. Il est donc capable de dire, en ayant simplement vu l'image réduite, si l'image est conforme à ses critères de recherche ou non. Pour d'autres images, leur contenu n'est pas visible (du fait de la forte décimation taille). Par conséquent, l'observateur est contraint de visualiser l'image originale afin de se faire une opinion. La figure 5 ((a) et (b)) reflète bien cette situation. Par exemple, le contenu de la dernière image de la première ligne, la première image de la seconde ligne ainsi que la troisième image de la dernière ligne de la figure 5 (a) est difficile à appréhender.

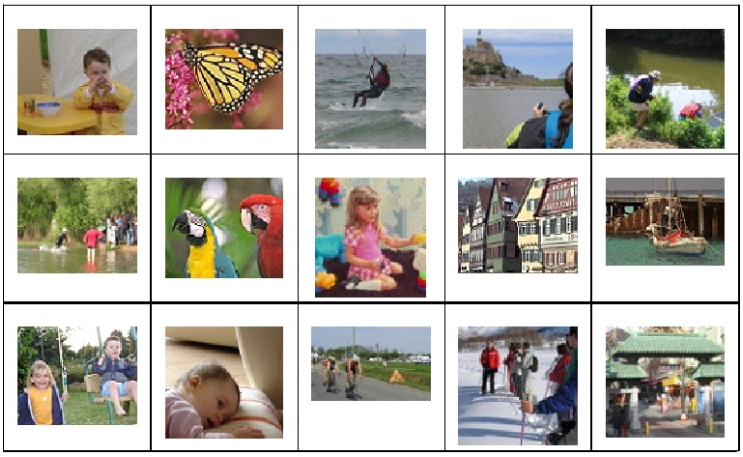
A partir d'une base de 50 images, présentant des contenus variés, une première évaluation qualitative a été réalisée. Du fait de la simplicité de l'évaluation effectuée, l'objectif est d'obtenir non pas une certitude mais une tendance.
Trois tests ont été effectués:

1. Test 1: les miniatures obtenues via une approche classique de décimation sont montrées à différents observateurs. Chacun d'eux doit dire rapidement pour chaque miniature si elle lui paraît exploitable ou pas. Le nombre de miniatures exploitables définit le score. Exploitable caractérise le fait qu'en un coup d'oeil l'utilisateur est capable de comprendre la scène rendue par la miniature ;

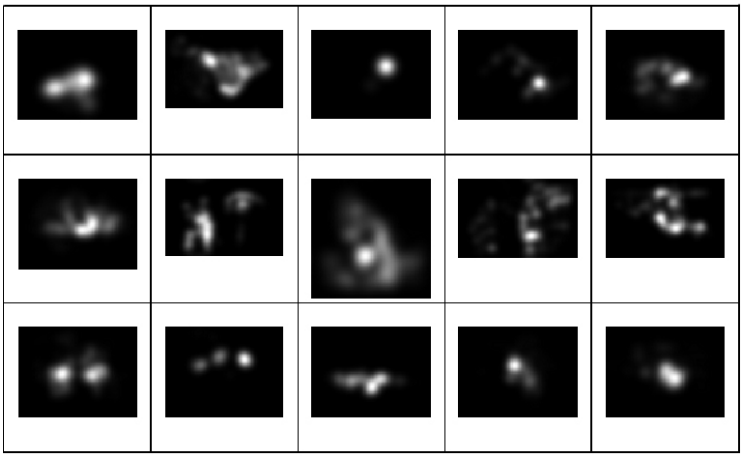
2. Test 2: les miniatures obtenues via les cartes de saillance sont montrées à différents observateurs. Comme précédemment, les observateurs doivent décider du caractère exploitable ou non de chaque miniature, d'où le score donné par le nombre de miniatures exploitables.

3. Test 3: ce dernier test est un test comparatif. Il s'agit de comparer les miniatures du test 1 avec celles du test 2 afin de déterminer qu'elle est la miniature la plus pertinente.

(a) Miniatures obtenues par une approche classique



(b) Miniatures obtenues par l'approche proposée



(c) Cartes de saillance associées

Fig. 5: Exemples d'images miniatures obtenues via une approche classique de décimation (a) et via la méthode proposée (b). Les cartes de saillance associées aux images (c) sont également données.

Tab. 3: Evaluation subjective de la pertinence de l'algorithme proposé et de l'algorithme de type classique: test 1 et test 2

Test 1 miniatures classique		Test 2 méthode proposée	
Exploitable	Inexploitable	Exploitable	Inexploitable
62%	38%	67%	33%

Le tableau 3 donne les résultats des tests 1 et 2. Sur une base de 50 images et pour 30 observateurs, le nombre de miniatures exploitable rapidement provenant du test 2 est supérieur au nombre de miniatures provenant du test 1: 67% des miniatures construites à partir de l'approche proposée sont jugées rapidement exploitables contre 62% des miniatures provenant de l'approche classique. La méthode proposée est donc légèrement plus performante que la méthode classique.

Tab. 4: Évaluation comparative des miniatures de l'algorithme proposé et de l'algorithme de type classique: test 3

Test 3	
Nombre de miniatures préférées provenant de l'algorithme classique	Nombre de miniatures préférées provenant de l'algorithme proposé
42%	58%

Le tableau 4 donne le résultat du test 3. On constate que les observateurs préfèrent les miniatures construites à partir des cartes de saillance (58% proviennent de la méthode proposée).
Finalement, la tendance que nous observons mets en évidence une préférence pour la miniaturisation des images basée sur une carte d'attention visuelle. Cette préférence reste faible et s'explique essentiellement par les raisons suivantes.
Tout d'abord, toutes les images de la base ont été réduites par la méthode proposée. Bien que l'algorithme soit piloté par le paramètre P permettant d'effectuer une sélection plus ou moins importante, la perte de contexte dans certain cas est très préjudiciable. L'approche systématique est donc à exclure. Ils auraient été certainement plus intéressant d'évaluer la pertinence de l'algorithme de décimation et de l'algorithme proposé afin de prendre une décision sur la stratégie à appliquer: décimation ou miniaturisation basée sur la carte d'attention ?
Par ailleurs, étant donné que le modèle d'attention visuelle n'inclue pas d'informations de haut niveau (détection de visage...), la pertinence de la carte d'attention visuelle est faible pour des images présentant des visages en gros plan. La conséquence directe est l'obtention d'une miniature présentant un problème de cadrage (un exemple est donné à la figure 6).
Enfin, le protocole de test en lui-même peut être remis en question. Un test comparatif avec référence est très pénalisant puisque le moindre défaut est sanctionné. Un test en aveugle aurait été certainement plus favorable. Enfin, le panel d'observateurs peut être remis en question. Pour cette première évaluation, les observateurs font tous parties de THOMSON R&D. La représentativité de l'échantillon n'est certainement pas la meilleure.
Ces interrogations, du fait du caractère exploratoire de ce travail, permetteront de définir un nouveau protocole de test. Ce dernier offrira la possibilité de montrer l'intérêt des vignettes obtenues via un modèle d'attention visuelle.

(a) Approche classique (b) Approche proposée

Fig. 6: Exemple d'un problème de re-cadrage: (a) approche classique (décimation); (b) approche proposée.

5 Conclusion

La saillance visuelle s'avère être une information très importante. Sa détermination, via une approche computationnelle fait l'objet de plus en plus d'études. Actuellement, les meilleures techniques de modélisation de l'attention visuelle font appels à des connaissances pluri-disciplinaires. Des données neurophysiologiques sont utilisées à la fois pour comprendre le fonctionnement global des mécanismes mis en jeu dans le cerveau et pour décliner les modèles algorithmiques biologiquement plausibles. Les expérimentations psychophysiques sont, quant à elle, requises pour identifier et évaluer les performances et les limitations de la perception visuelle.
Dans cette contribution, un modèle construisant une carte de saillance finale à partir de trois cartes de saillance est proposé. Une nouvelle évaluation de performances est réalisée. Cette nouvelle approche de modélisation de l'attention visuelle est intégrée dans un système de vision destiné à réaliser des recadrages et des zooms automatiques centrés sur les régions perceptuellement importantes. L'application directe concerne l'adaptation des contenus pour des dispositifs d'affichage à écrans de petites tailles.
La première évaluation qualitative de ce système de vision est encourageante puisque les résultats obtenus par l'approche proposée sont considérés meilleurs que ceux de l'approche classique. Ces résultats sont d'autant plus intéressants qu'il existe un fort potentiel d'amélioration de ce système de vision. Tout d'abord, l'approche computationnelle de modélisation de l'attention visuelle est purement basée sur des données de bas niveaux. Aucun aspect sémantique et cognitif n'est actuellement considéré. Par ailleurs, le procédé de miniaturisation est actuellement systématique alors que toutes les images ne peuvent pas être réduites. L'algorithme doit donc être adapté. Une analyse sémiologique peut permettre de donner des indications sur le degré d'acceptabilité des observateurs selon les types de contenu.
Les perspectives sont également très prometteuses puisque l'extension de ce système de vision à la vidéo est en cours de réalisation. Cependant la prise en compte de l'aspect temporelle fait apparaître d'autres problèmes tels que la cohérence temporelle de recadrage des images réduites.

References

[Fan 03] X. Fan, X. Xie, W.Y. Ma, H.J. Zhang, H.Q. Zhou. – Visual attention based image browsing on mobile devices. – *in Proc. of ICME 2003*, vol. 1, 2003.

[Itti 05] L. Itti. – Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12(6):1093–1123, Aug 2005.

[Koch 85] C. Koch, S. Ullman. – Shifts in selection in visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–227, 1985.

[Krauskopf 82] J. Krauskopf, D. R. Williams, D. W. Heeley. – Cardinal direction of color space. *Vision Research*, 22:1123–1131, 1982.

[Landragin 04] F. Landragin. – Saillance physique et saillance cognitive. *Cognition, Représentation, Langage*, 2(2), 2004.

[Le Callet 99] P. Le Callet, A. Saadane, D. Barba. – Orientation selectivity of opponent-colour channels. – *ECVP'99*, 1999.

[Le Meur 05] O. Le Meur. – *Attention sélective en visualisation d'images fixes et animées affichées sur un écran: modèles et évaluation de performances. Applications.* – Université de Nantes, PhD. Thesis, Ecole Polytechnique de l'Université de Nantes, IRCCyN, 2005.

[Le Meur 06] O. Le Meur, P. Le Callet, D. Barba, D. Thoreau. – A coherent computational approach to model bottom-up visual attention. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(5):802–818, 2006.

[Liu 03] H. Liu, X. Xie, W.Y. Ma, H.J. Zhang. – Automatic browsing of large pictures on mobile devices. – *in ACM Multimedia conference*, 2003.

[Senane 93] H. Senane, A. Saadane, D. Barba. – The computation of visual bandwiths and their impact in image decomposition and coding. – *Proceedings HVEI-93*, 1993.

[Watson 87] A. B. Watson. – The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics and Image Processing*, 39:311–327, 1987.

[Wolfe 04] J.M. Wolfe, T.S. Horowitz. – What attributes guide the deployement of visual attention and how do they do it. *Nature*, 5, 2004.